Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 115 (2017) 307–311

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

Multiple Classification Using SVM Based Multi Knowledge Based System

Thirumalaimuthu Thirumalaiappan Ramanathan*, Dharmendra Sharma

Faculty of Education, Science, Technology and Mathematics, University of Canberra, Australia

Abstract

Support Vector Machine (SVM) is a machine learning classification technique that supports binary classification. In the recent years, efforts are made to extend the SVM algorithm to support multiple classifications. This paper presents a SVM based multi-knowledge-based system (SMK) design that supports multiple classifications. The proposed design is successfully tested on a classification problem. The benchmark car evaluation dataset from UCI machine learning repository is used for training and testing the SMK. The SMK shows good performance on this classification and shows good promise for the future.

*Keywords:*Support vector machine, knowledge-based system;

* Thirumalaimuthu Thirumalaiappan Ramanathan. Tel.: +61-469703287.
E-mail address: thirumalaimuthu.thirumalaiappanramanathan@canberra.edu.au

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications
10.1016/j.procs.2017.09.139

1. Introduction

A support vector, multi-knowledge based system design (SMK) proposed in this paper shows a multiple classification approach. SMK is designed to support classification through extracting patterns from datasets. In the SMK design (Figure 1), the dataset is divided into multiple groups on the basis of output class. Suppose that the dataset contains four output class, the dataset is divided into three groups, where the first group contains data that falls under the first and second output classes, the second group contains data that falls under the second and third output classes and the third group contains data that falls under the second and third output classes. SVM [1] is used to extract patterns from the divided datasets groups. The patterns extracted from different datasets groups are used in terms of rules in different knowledge bases systems (KBS) [2]. The SMK connects these different KBS through a main KBS to support multiple classifications.

In the SMK design, the main KBS checks the output of KBS which contains the rules that classify the data into first and second class, if output is the second class, the main KBS checks output of the KBS which contains the rules that classify the data into second and third class, and if the output is third class, the main KBS finally checks the output of the KBS which contains rules that classify the data into third and fourth class and deliver the final output class. Below gives small introduction about SVM, KBS before describing the SMK.

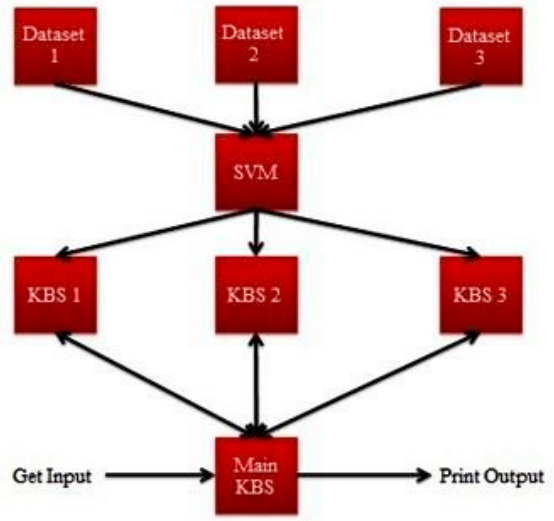

Fig.1. SMK Architecture.

2. SVM

SVM is primarily a classier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels. SVM consists of a learning module (svm-learn) and a classification module (svm-classify). The training model takes the input file, target file and trains the network. In the classification model, the various class labels like class l, 2, and 3 ....., 247 are given. Thus the SVM learns and produces correct labels of the classes [3]. For the SMK, SVM is used to extract patterns from the datasets.

3. KBS

KBS is a computer program and uses a knowledge base to solve complex problems. KBS was first developed by artificial intelligence researchers. It is essentially composed of two sub-systems: the knowledge base and the inference engine. The knowledge base represents facts about the world. The inference engine is an automated reasoning system that evaluates the current state of the knowledge-base, applies relevant rules, and then asserts new knowledge into the knowledge base [4]. Forward chain KBS method is used to build the SMK.

4. Dataset

Car evaluation dataset used for the classification is a benchmark dataset from [3]. It contains 1728 instances. The input attributes are buying price, maintenance price, number of doors, capacity in terms of persons to carry, size of luggage boot, estimated safety of the car. The output classes are unacceptable, acceptable, good, very-good. For the SMK, only three inputs: buying price, maintenance price and estimated safety are used. SVM deals with binary inputs, so the available six inputs have to be grouped into possible binary inputs which will be more complex. So only three inputs are chosen for the SMK which will be simple for grouping and training the SVM, also the main focus is to investigate whether the SVM can support multiple classifications through KBS.
The dataset is grouped into three groups of datasets where the first group contains data that falls under the categories "unacceptable" and "acceptable", the second group contains data that falls under the categories "acceptable" and "good" and the third group contains data that falls under the categories "good" and "very good".
Out of 1728 instances from the dataset, the first 1678 instances are used to train the SVM and the remaining 50 instances are used to test the SMK design.

Table 1. Sample Dataset [3].

Buying Price	Maintenance Price	Estimated Price	Output
Low	Med	High	Good
Med	High	Low	Acceptable
Vhigh	Vhigh	Med	Unacceptable

Table 1 shows sample data from the car evaluation dataset. The inputs buying price and maintenance price have four possible data: low, medium, high and very high. The input estimated safety has three possible data: low, medium and high. The output is either unacceptable, acceptable, good or very good.
Using this dataset, the main research question that is asked from the proposed SMK design is

- Can a SVM algorithm be extended to use in a multiple KBS to support multiple classification?
- Can such a system design show good classification accuracy?

5. SMK Algorithm

Below is the SMK algorithm.
Step 1. Get input.
Step 2. Divide dataset based on output class
Step 3. SMK()
First KBS ()
Train SVM using first group dataset ();
For all possible combination of inputs
SVM Classification ();
Add SVM Results to rules ();
Second KBS ()
Train SVM using second group dataset ();
For all possible combination of inputs
SVM Classification ();
Add SVM Results to rules ();
Third KBS ()
Train SVM using third group dataset ();
For all possible combination of inputs
SVM Classification ();
Add SVM Results to rules ();
Main KBS ()
First KBS ();

If (output (First KBS) == "unacceptable")
then Print output (First KBS);
Else Second KBS ();
If (output (Second KBS) == "acceptable")
Then Print output (Second KBS);
Else Print output (Third KBS);

As described in the above steps, the multiple output class datasets is divided into different dataset groups based on the output class. For the car evaluation dataset which has four output classes, the dataset is divided into three groups. SVM is used to train these different groups of dataset. Then the SMK design checks the classification results from the trained SVM for all possible combination of inputs. Then these results are stored in terms of if-then rules in different KBS such that the rules containing the SVM classification results from pattern extraction of first group dataset is stored in first KBS and so on. Likewise three KBS are formed for storing the rules extracted from three groups of datasets respectively. The main KBS in SMK design checks and compare the output of different KBS as shown in the above algorithm and prints the result of the output class. MATLAB [6] is used for implementing the SMK algorithm.

6. Implementation

MATLAB is used for training the SVM using different datasets and SVM classification is tested for all possible combination of binary inputs. Figure 2 shows the SVM training from different datasets.

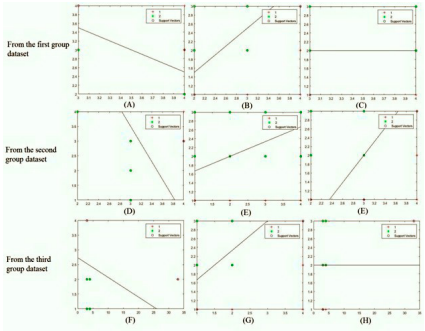

Fig.2. SVM training for inputs (a, d, e) buying price and maintaining price, (b, e, g) maintaining price and estimated safety, (c, e, h) buying price and estimated safety under the output class

In the Figure 2, the first row shows SVM training for three possible combinations of inputs using the first dataset. The second, third rows shows SVM training from the other two datasets which are grouped from the original dataset for different datasets training as described in the section 4. The results from the SVM training using different datasets are stored separately. These results are used in terms of rules using a multiple KBS. Three KBS is designed as three types of class that uses forward chaining approach in terms of if-then rules. The results from the SVM training of the first group dataset are stored in one KBS, and similarly the other two KBS contains the results extracted from the other two groups of datasets. The main KBS is designed to check and compare the results of different KBS and displays the results.

7. Performance of SMK

The performance of SMK is estimated using classification accuracy. The classification accuracy is calculated using true positive (TP), false positive (FP) [8] and using true negative (TN), false negative (FN) [9].

Classification Accuracy = (TP+TN) / (TP+TN+FP+FN).

TP: These are the positive tuples that were correctly labeled by the classifier. If outcome from a prediction is p and the actual value is also p, then it is called TP.

TN: These are the negative tuples that were correctly labeled by the classifier.

FP: These are the negative tuples that were incorrectly labeled by the classifier. However if the actual value is in then it is said to be false positive.

FN: These are the positive tuples that were mislabeled as negative.

The SMK is tested using the testing dataset (Section 2) which gave a classification accuracy of 98%. When testing, out of 50 input data, 49 showed correct classifications resulted from the system. Each of three knowledge base system showed similar classification accuracy similar to the final output from the program.

8. Conclusion and Future Work

SMK design shows 98% of accuracy in classification. SVM algorithm supports binary classification. But through SMK design proposed in this paper, the SVM shows four types of classification by using multiple KBS. Although previously various techniques were proposed to support SVM multi-class classification such as one-against-all method [10], one-against-one method [11], Directed Acyclic Graph SVM [12] and extreme learning machine [13], these techniques lacks in handling large amounts of unstructured data in an intelligent fashion. To overcome these drawbacks, SMK design is proposed in this paper that uses SVM in multiple knowledge based system to handle large amounts of unstructured data and to support multi-class classification. The future work will be improving SMK design to handle incomplete data by using fuzzy logic.

References

[1] C. Cortes, and V. Vapnik. Support vector machine. *Machine learning* 1995; **20**(**3**): 273-297.
[2] R.G Smith. Knowledge-based systems: Concepts, techniques, examples. *Canadian High Technology Show* 1985.
[3] R. Seethalakshmi, T.R. Sreeranjani, T. Balachandar, A. Singh, M. Singh, R. Ratan and S. Kumar. 2005. Optical character recognition for printed Tamil text using Unicode. *Journal of Zhejiang University Science* 2005; **6**(**11**):1297-1305.
[4] F. Hayes-Roth, D. Waterman and D. Lenat. Building expert systems. *Addison-Wesley* 1984; ISBN 0-201-10686-8.
[5] M. Bohanec and B. Zupan. UCI Machine Learning Repository 1997. "https://archive. ics. uci. edu/ml/datasets/Car+Evaluation".
[6] M.A. Natick. M.U.S. Guide. *The mathworks. Inc.* 1998; 5: 333.
[7] A. Hejlsberg, S. Wiltamuth and P. Golde. C# language specification. *Addison-Wesley Longman Publishing Co., Inc*. 2003.
[8] M. Sapna Jain, A. Afshar and M.N. Doja. K-Means clustering using Weka Interface. *Proceedings of the 4th National Conference*. *Computing For Nation Development* 2010.
[9] J. Han, M. Kamber, and J. Pei. Data mining: concepts and techniques. *Elsevier* 2011.
[10] L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. A. Muller, E. Sackinger, P. Simard, and V. Vapnik. Comparison of classifier methods: a case study in handwriting digit recognition. *In International Conference on Pattern Recognition* 1994; 77–87.
[11] J. H. Friedman. Another approach to polychotomous classification. *Technical report, Department of Statistics, Stanford University* 1996.
[12] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin DAGs for multiclass classification. *In Advances in Neural Information Processing Systems MIT Press* 2000; **12**: 547–553.
[13] G.B. Huang, H. Zhou, X. Ding and R. Zhang. 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics*) 2012; **42**(**2**): 513-529.